

WanaIQ

Technical Roadmap for MVP Development

AI-Powered Civic Engagement Platform for Democratic Participation

Project Name:	WanaIQ: AI-Powered Civic Engagement Platform
Thematic Area:	Generative & Agentic AI
Document Version:	1.5 (Final Submission)
Submission Date:	January 30, 2026
Development Period:	February 2026 - March 2026 (8-Week Sprint)
Development Approach:	Vertical Slice Architecture - Solo Developer
Current Status:	Ready for Incubation Phase

■ Key Highlights

- ✓ **Resource-Efficient AI:** Hybrid inference strategy eliminates 24/7 heavy model costs through on-demand serverless architecture
- ✓ **Kenyan Context Expertise:** Custom Swahili NLP models and 500+ curated Kenyan civic documents
- ✓ **Agentic AI Innovation:** Autonomous 'Civic Agent' that routes issues and guides citizens through complex processes
- ✓ **Mobile-First Design:** Progressive Web App optimized for Kenya's mobile-first, bandwidth-constrained market
- ✓ **Proven Velocity:** Vertical slice architecture enables rapid iteration with demo-ready features every week

1. Executive Summary

This technical roadmap outlines the **expedited 8-week development plan** for WanaIQ's Minimum Viable Product (MVP). Designed for sprint development ending in **March 2026**, this plan leverages high-efficiency tools (React, Node.js, Supabase, Agentic AI) to deliver a production-grade civic engagement solution.

By employing a **Vertical Slice Architecture** as a solo developer assuming multiple functional roles, WanaIQ will deploy fully functional features incrementally, ensuring a **demo-ready product at every milestone**. This approach maximizes development velocity while maintaining code quality and user experience standards.

1.1 MVP Core Objectives

Objective	Description	Success Metric
■ Agentic AI Assistant	Deploy 'Civic Agent' that intelligently routes citizen issues using LLM-driven classification and multi-step workflow execution	<ul style="list-style-type: none">• 90%+ routing accuracy• <2s response time• 200+ queries/day
■ RAG Pipeline ("Civic Brain")	Enable citizens to query Kenyan Constitution and laws in natural language with verified, cited responses.	<ul style="list-style-type: none">• 85%+ answer accuracy• 100% source citation• 10+ languages supported
■ Content Analysis AI	Automated video transcription, sentiment analysis, and metadata extraction for CivicClips educational content	<ul style="list-style-type: none">• <2 min processing time• 95%+ transcription accuracy• Auto-moderation active
■ Cloud-Native Scalability	Serverless architecture using Next.js, Node.js, and Supabase for high performance and auto-scaling.	<ul style="list-style-type: none">• <200ms API latency (p95)• 99.9%+ uptime• 1000+ concurrent users
■ Multilingual Support	Real-time translation supporting English, Swahili, and 10+ Kenyan languages for inclusive civic access	<ul style="list-style-type: none">• 10+ languages live• <500ms translation time• 80%+ quality score

2. Technical Architecture Overview

2.1 Technology Stack

WanaIQ employs a modern, production-grade technology stack optimized for rapid development, resource efficiency, and scalability. Each technology choice is justified by specific technical requirements:

Component	Technology	Purpose	Justification
Frontend Web	React 18 + TypeScript Next.js 14	Responsive web with SSR/SSG	SEO optimization, fast load, type safety
Frontend Mobile	React Native + Expo	Cross-platform iOS/Android	Code reuse 70%+, rapid iteration
Backend API	Python/Node.js + Express GraphQL	RESTful and GraphQL APIs	Async I/O for AI calls, reduces over-fetching
Database (Primary)	Supabase PostgreSQL + pgvector	Structured data + vectors	Built-in auth, real-time, vector search
Database (Content)	MongoDB Atlas	Unstructured content	Flexible schema, excellent performance
AI/ML Core	Llama 3 via Groq API	LLM inference	Sub-100ms latency, 10x faster, cost-effective
Vector Database	Supabase pgvector	RAG embeddings	Native PostgreSQL, no separate service
NLP Processing	spaCy + Custom models	Swahili/English NLP	Production-ready, custom Kenyan training
Computer Vision	OpenCV + Python	Image verification	Microservice, on-demand scaling
Transcription	Whisper API	Video transcription	State-of-art accuracy, multi-language
Caching	Redis Cloud	Session + response cache	70% cost reduction, sub-ms latency
Cloud Infrastructure	AWS Lambda + EC2 + S3 / Supabase + S3	Serverless + hosting	Pay-per-use, auto-scaling, global CDN
CDN	Cloudflare	Content delivery	DDoS protection, edge caching, analytics
Authentication	OAuth 2.0 + JWT	Secure auth	Industry standard, social login, stateless
CI/CD	Vercel + GitHub Actions	Automated deployment	Zero-config deploys, auto previews, rollback

2.2 System Architecture

WanaIQ follows a **microservices architecture** with clear separation of concerns, enabling independent scaling and deployment of each component:

Layer	Components	Responsibilities	Scaling Strategy
1. Presentation Layer	<ul style="list-style-type: none">• Next.js web app• React Native mobile• PWA shell	<ul style="list-style-type: none">• Server-side rendering• Static generation• Client-side hydration• Responsive design	Edge caching via CDN, static asset optimization, code splitting
2. API Gateway Layer	<ul style="list-style-type: none">• Express.js REST API• GraphQL endpoint• WebSocket server	<ul style="list-style-type: none">• Request routing• Authentication/authorization• Rate limiting• Response caching	Horizontal scaling via load balancer, stateless design
3. Application Logic Layer	<ul style="list-style-type: none">• Business logic• Data validation• Workflow orchestration	<ul style="list-style-type: none">• User management• Content moderation• Notification system• Analytics processing	Containerized services, auto-scaling based on CPU/memory
4. AI Services Layer	<ul style="list-style-type: none">• Civic Agent Router• RAG Engine• Video Analyzer• Translation service	<ul style="list-style-type: none">• LLM inference• Vector search• CV processing• NLP analysis	Serverless functions, on-demand execution, cold start optimization
5. Data Layer	<ul style="list-style-type: none">• PostgreSQL• MongoDB• Redis• S3 storage	<ul style="list-style-type: none">• Persistent storage• Document storage• Cache management• Media storage	Database replication, read replicas, sharding ready

3. Development Phases & Milestones (8-Week Sprint)

The 8-week development sprint is organized into three progressive phases, with each milestone delivering testable, demo-ready functionality:

Milestone	Timeline	Deliverables	Success Criteria	Testing
M1.1 Infrastructure Setup	Feb Week 1	<ul style="list-style-type: none">• Git linked to portal• Supabase configured• CI/CD pipeline• Development environment• Staging deployment	<ul style="list-style-type: none">✓ Repo visible in portal✓ Database accepts connections✓ Auto-deploy on commit✓ 100% env documented	<ul style="list-style-type: none">• Connection tests• Pipeline smoke tests• Access control verification
M1.2 API Migration	Feb Week 1-2	<ul style="list-style-type: none">• Node.js Express server• GraphQL schema• Authentication middleware• Rate limiting• API documentation	<ul style="list-style-type: none">✓ All endpoints responding✓ Auth flow working✓ 100 req/min limit enforced✓ Swagger docs live	<ul style="list-style-type: none">• Unit tests (80% coverage)• Integration tests• Load testing (100 users)
M2.1 Civic Agent Router	Feb Week 3	<ul style="list-style-type: none">• LLM classification logic• Department routing• Intent recognition• Multi-turn conversation• Confidence scoring	<ul style="list-style-type: none">✓ 90%+ routing accuracy✓ <2s response time✓ 5+ issue categories✓ Conversation state persists	<ul style="list-style-type: none">• 100 test cases• Accuracy benchmarking• Latency profiling• Edge case testing
M2.2 RAG Pipeline	Feb Week 4	<ul style="list-style-type: none">• Document ingestion (500+ docs)• Vector embeddings• Similarity search API• Context injection• Source citations	<ul style="list-style-type: none">✓ 85%+ answer accuracy✓ 100% responses cited✓ <1s search latency✓ No hallucinations detected	<ul style="list-style-type: none">• 50 civic Q&A pairs• Citation validation• Hallucination detection• Performance testing
M2.3 CivicClips AI	Mar Week 1	<ul style="list-style-type: none">• Video upload pipeline• Whisper transcription• Sentiment analysis• Auto-captioning• Content moderation	<ul style="list-style-type: none">✓ <2 min processing✓ 95%+ transcription accuracy✓ Sentiment scored✓ Inappropriate content flagged	<ul style="list-style-type: none">• 20 video samples• Accuracy validation• Processing time tests• Moderation precision
M3.1 Gamification System	Mar Week 2	<ul style="list-style-type: none">• Civic points logic• User reputation scores• Achievement badges• Leaderboards• Quest system	<ul style="list-style-type: none">✓ Points awarded correctly✓ Leaderboard updates real-time✓ 10+ achievements✓ Engagement +50%	<ul style="list-style-type: none">• Points calculation tests• Leaderboard accuracy• Performance under load• User acceptance testing
M3.2 Mobile Optimization	Mar Week 2-3	<ul style="list-style-type: none">• Progressive Web App• Offline capability• Push notifications• Touch gestures• App manifest	<ul style="list-style-type: none">✓ Lighthouse score >90✓ Offline mode functional✓ Notifications working✓ iOS/Android tested	<ul style="list-style-type: none">• Device compatibility• Offline scenario tests• Performance audit• User experience testing
M3.3 Final Launch	Mar Week 3-4	<ul style="list-style-type: none">• Production deployment• SSL certificates• Monitoring setup• Demo video• Documentation	<ul style="list-style-type: none">✓ Public URL live✓ 99.9%+ uptime✓ Demo video <5 min✓ All docs complete	<ul style="list-style-type: none">• Full regression testing• Security audit• Performance benchmarking• User acceptance testing

4. Development Strategy: Vertical Slice Architecture

As a **solo developer** participating in this hackathon, the project employs a Vertical Slice Architecture where I assume multiple 'Functional Hats' to execute the roadmap efficiently. This approach, augmented by AI tools and modern frameworks, enables single-developer velocity comparable to a small team:

Functional Hat	Core Responsibilities	Efficiency Tools ("Virtual Team")	Time Allocation
■■■ Full Stack Lead	<ul style="list-style-type: none">• React/Node.js development• Database schema design• API security implementation• Code review and refactoring	Supabase: Instant backend (auth, DB, real-time) Vercel: Zero-config CI/CD shadcn/ui: Pre-built components Copilot: Code completion	40% of time (~16 hrs/week)
■ AI/ML Engineer	<ul style="list-style-type: none">• RAG pipeline construction• LLM prompt engineering• Agentic workflow logic• Model evaluation	LangChain: Agent orchestration HuggingFace: Pre-trained models Groq API: Fast LLM inference Pinecone: Vector database	30% of time (~12 hrs/week)
■ Product & Design	<ul style="list-style-type: none">• UI/UX implementation• User story mapping• Feature prioritization• Demo creation	Tailwind CSS: Rapid styling v0.dev: AI UI generation Figma: Design mockups Loom: Demo videos	20% of time (~8 hrs/week)
■ DevOps & QA	<ul style="list-style-type: none">• Infrastructure setup• Monitoring configuration• Test automation• Performance optimization	GitHub Actions: CI/CD Playwright: E2E testing Datadog: Monitoring Sentry: Error tracking	10% of time (~4 hrs/week)

Key Productivity Multipliers:

- **AI-Assisted Development:** GitHub Copilot, Claude, and ChatGPT reduce coding time by 40-50% for boilerplate and standard patterns
- **Managed Services:** Supabase, Vercel, and MongoDB Atlas eliminate infrastructure management overhead
- **Component Libraries:** shadcn/ui and Tailwind provide production-ready components, avoiding custom CSS
- **Testing Automation:** Playwright and Jest enable continuous testing without manual QA efforts
- **Low-Code Tools:** v0.dev generates initial UI layouts, reducing design-to-code translation time by 60%

5. AI/ML Technical Approach (Resource-Efficient Strategy)

5.1 Hybrid Inference Architecture

To avoid the 'resource-heavy' trap of running large AI models 24/7, WanaIQ implements a strategic **hybrid inference approach** that separates heavy and light workloads:

Workload Type	Processing Method	Infrastructure	Cost Model	Latency
■ Heavy Lifting (Video/Image)	On-Demand Serverless Functions	AWS Lambda with Python runtime (2GB memory)	Pay-per-invocation \$0.20/100 videos vs \$50/mo always-on	30-120s (acceptable for async tasks)
■ Agent Logic (Text/Chat)	API-Based Inference (Groq/Replicate)	Managed API endpoints with global CDN	Pay-per-token \$0.10/1M tokens vs \$200/mo GPU	<100ms (critical for real-time chat)
■ Vector Search (RAG)	Managed Vector DB (Supabase pgvector)	PostgreSQL extension with indexing	Included in DB costs \$25/mo database	<50ms (optimized with indexes)
■ Batch Processing (Analytics)	Scheduled Jobs (Cron + Lambda)	Serverless functions run off-peak hours	\$0.50/day for daily jobs	Not time-critical (overnight OK)

5.2 RAG over Fine-Tuning: Strategic Decision

Factor	RAG Approach (WanaIQ Choice)	Fine-Tuning Approach	Decision
Development Time	1-2 weeks to implement with existing documents	4-6 weeks for data prep, training, and validation	✓ RAG 3x faster
Infrastructure Cost	\$0/month (API + vector DB)	\$500-1000/month (GPU training + serving)	✓ RAG is cheaper
Update Flexibility	Add new documents in minutes, instant availability	Retrain model (days), validate, redeploy	✓ RAG real-time updates
Accuracy & Verification	Responses grounded in actual docs, 100% source citations	Model may hallucinate, difficult to verify sources	✓ RAG verifiable
Hackathon Suitability	MVP ready in 1 week, demo-friendly	Requires weeks before usable results	✓ RAG sprint-compatible

Conclusion: RAG provides superior accuracy, flexibility, and development speed. Fine-tuning requires expensive hardware. RAG (Retrieval-Augmented Generation) costs virtually nothing on our stack because we store embedding's in our existing PostgreSQL database and inject context into the LLM prompt. This is not just cheaper, it's more accurate for legal documents

5.3 Detailed AI Component Specifications

Component	Technical Specification	Performance Target	Fallback Strategy
Civic Agent Router	<ul style="list-style-type: none">Model: Llama 38B via GroqContext: 8k tokensTemperature: 0.3Few-shot: 5 examples/category	<ul style="list-style-type: none">Accuracy: >90%Latency: <2sThroughput: 100 req/min	Fallback to rule-based classifier if API down; alert admin
RAG Engine	<ul style="list-style-type: none">Embeddings: sentence-transformersVector DB: pgvectorChunk size: 512 tokensTop-k retrieval: 5 chunks	<ul style="list-style-type: none">Answer accuracy: >85%Search latency: <500msCitation rate: 100%	Serve cached responses; queue queries for batch processing
Video Transcription	<ul style="list-style-type: none">Model: Whisper Large v3Languages: EN, SW, 10+Format: SRT captionsPost-process: Punctuation, capitalization	<ul style="list-style-type: none">Accuracy: >95% WERProcessing: <2 min/videoSupported: 15+ languages	Use Whisper Medium for faster processing; manual correction UI
Sentiment Analysis	<ul style="list-style-type: none">Model: Custom spaCy pipelineTraining: Kenyan social media corpusClasses: Positive, Negative, NeutralConfidence threshold: 0.7	<ul style="list-style-type: none">F1 score: >0.80Latency: <100msBatch: 1000 posts/sec	Generic sentiment model if custom unavailable; manual review flag
Translation	<ul style="list-style-type: none">Service: Google Translate APILanguages: 15+ Kenyan languagesCaching: Redis (24h TTL)Quality: Human validation sample	<ul style="list-style-type: none">Cache hit rate: >70%Latency: <300msCost: <\$10/day	Show original with translation disclaimer; community corrections

6. Risk Management & Mitigation Strategies

Risk Category	Impact Level	Potential Impact	Mitigation Strategy	Contingency Plan
■ Technical: Integration Complexity	MEDIUM	Node.js + Python services fail to communicate, blocking AI features	<ul style="list-style-type: none">• Strict API contracts (OpenAPI)• Redis message queues• Comprehensive integration tests• Mock services for development	<ul style="list-style-type: none">• Fallback to monolithic architecture temporarily• Manual data transfer• Extend timeline by 1 week
■ AI/ML: Model Hallucinations	HIGH	AI provides incorrect legal information, damaging user trust and platform credibility	<ul style="list-style-type: none">• Strict RAG grounding prompts• Response validation layer• Confidence thresholds (>0.8)• Human review for low-confidence• Explicit disclaimers	<ul style="list-style-type: none">• Disable AI chat temporarily• Switch to curated FAQ• Community fact-checking• Legal expert hotline
■ Security: Data Privacy Breach	CRITICAL	User data exposed, legal liability, reputational damage, user exodus	<ul style="list-style-type: none">• Encryption at rest and transit• Regular security audits• Penetration testing (week 6)• Minimal data collection• Access controls + audit logs	<ul style="list-style-type: none">• Incident response plan• Immediate user notification• Data breach insurance• Legal compliance team• Temporary shutdown if needed
■ Operational: API Rate Limits	MEDIUM	LLM API throttling during demos, poor user experience, high costs	<ul style="list-style-type: none">• Request queuing system• Load balancing across providers• Aggressive caching (Redis)• Rate limit monitoring• Budget alerts (\$100/day cap)	<ul style="list-style-type: none">• Fallback to lighter models• Queue non-urgent requests• Upgrade API tier• Implement waiting room
■ Content: Multilingual Support	MEDIUM	Poor Swahili translation quality, alienates majority users	<ul style="list-style-type: none">• Custom spaCy models for Swahili• Community validation program• Human translators for critical content• Progressive language expansion• User feedback loops	<ul style="list-style-type: none">• English-only MVP first• Community-driven translations• Partner with translation NGOs• Delay language launch by 2 weeks
■ Timeline: Scope Creep	HIGH	Feature additions delay MVP launch, miss March deadline	<ul style="list-style-type: none">• Strict milestone-based development• Weekly scope reviews with mentor• Feature flags for optional items• "Must have" vs "Nice to have" matrix• Time-boxed sprints	<ul style="list-style-type: none">• Cut gamification (M3.1)• Reduce language support to 3• Skip mobile optimization• Focus on core AI features only

7. Success Metrics & Key Performance Indicators

WanaIQ's success will be measured across four dimensions: **technical performance**, **user engagement**, **AI effectiveness**, and **civic impact**. These metrics guide development priorities and validate our approach:

Category	Metric	MVP Target	Measurement Method	Review Frequency
■ Technical Performance	API Response Time (p95)	< 200ms	Datadog APM monitoring	Real-time dashboard
	Page Load Time (LCP)	< 2.5s	Lighthouse CI, Web Vitals	Every deployment
	System Uptime	> 99.9%	UptimeRobot, Pingdom	Daily report
	Error Rate	< 0.1%	Sentry error tracking	Real-time alerts
	Test Coverage	> 80%	Jest coverage reports	Every commit (CI)
■ AI Effectiveness	Agent Routing Accuracy	> 90%	Ground truth validation (100 cases)	Weekly evaluation
	RAG Answer Accuracy	> 85%	Human evaluation (50 Q&A pairs)	Weekly evaluation
	Hallucination Rate	< 5%	Manual review + user reports	Daily sampling (20 responses)
	AI Cache Hit Rate	> 70%	Redis analytics	Daily report
	Transcription Accuracy (WER)	< 5%	Manual validation (10 videos/week)	Weekly evaluation
■ User Engagement	Monthly Active Users	500+	Mixpanel analytics	Weekly cohort analysis
	Daily Active Users	100+	Session tracking	Daily dashboard
	Avg. Session Duration	> 5 minutes	Google Analytics	Weekly report
	7-Day User Retention	> 30%	Cohort retention analysis	Weekly cohort tracking
	AI Assistant Usage	200+ queries/day	Database query logs	Daily metrics
■■ Civic Impact	Civic Reports Submitted	50+	Action Hub database	Weekly tally
	CivicClips Views	1000+	Video analytics	Weekly report
	Promise Tracking Votes	500+	Voting system metrics	Weekly tally
	Languages Supported	10+	Translation system config	Milestone-based
	User Satisfaction (NPS)	> 40	In-app surveys (monthly)	Monthly survey

8. Visual Development Timeline (Gantt Chart)

Phase/Milestone	Feb W1	Feb W2	Feb W3	Feb W4	Mar W1	Mar W2	Mar W3	Mar W4
PHASE 1: Foundation	■■■	■■■						
M1.1: Infrastructure	■■■							
M1.2: API Migration	■■■	■■■						
PHASE 2: AI Core			■■■	■■■	■■■			
M2.1: Civic Agent			■■■					
M2.2: RAG Pipeline				■■■				
M2.3: CivicClips AI					■■■			
PHASE 3: Production						■■■	■■■	■■■
M3.1: Gamification						■■■		
M3.2: Mobile PWA						■■■	■■■	
M3.3: Launch & Demo							■■■	■■■
CONTINUOUS ACTIVITIES								
Testing & QA	■■	■■	■■	■■	■■	■■	■■	■■
Documentation	■■	■■	■■	■■	■■	■■	■■	■■
Mentor Check-ins	✓	✓	✓	✓	✓	✓	✓	✓

Legend: ■■■ = Active development | ■■ = Continuous activity | ✓ = Scheduled checkpoint

9. Conclusion & Commitment to Excellence

This technical roadmap provides a comprehensive, actionable blueprint for delivering WanaIQ's MVP within the **8-week sprint timeline ending in March 2026**. Our approach balances technical innovation with pragmatic resource management, ensuring we build a functional, demonstrable platform that showcases the transformative potential of Generative and Agentic AI in civic engagement.

■ Key Differentiators

- ✓ **Resource-Efficient AI:** Hybrid inference strategy eliminates 24/7 heavy model costs through serverless architecture, reducing operational costs by 90% while maintaining performance
- ✓ **Kenyan Context Expertise:** Custom Swahili NLP models and 500+ curated civic documents ensure culturally relevant, accurate responses that generic AI platforms cannot match
- ✓ **Mobile-First Design:** Progressive Web App optimized for Kenya's mobile-first, bandwidth-constrained market with offline capabilities and sub-2s load times
- ✓ **Agentic AI Innovation:** Goes beyond simple chatbots to implement autonomous agents that actively guide citizens through multi-step civic processes, demonstrating true agentic capabilities
- ✓ **Proven Solo Velocity:** Vertical slice architecture combined with AI-assisted development tools enables single-developer productivity comparable to a 3-4 person team

■ Success Validation

Each milestone includes **quantifiable success criteria** measured through automated testing, human evaluation, and user feedback. Weekly progress updates will demonstrate milestone completion with working demos, ensuring continuous validation of our approach.

■ Mentor Engagement

The WanaIQ development process is **committed to transparent execution**. Our Git repository is linked to the hackathon portal providing real-time visibility into development progress. We welcome mentor guidance throughout the incubation journey and commit to:

- Weekly progress updates with working demo videos showcasing completed milestones
- Bi-weekly live demo sessions for feedback and course correction
- Real-time code visibility through public GitHub repository with detailed commit messages
- Documented decision rationale for all major technical choices
- Proactive communication of blockers or timeline risks within 24 hours of identification

■ Vision Beyond MVP

While this roadmap focuses on MVP delivery, WanaIQ's architecture is designed for national scale from day one. Post-hackathon growth will expand to all 47 Kenyan counties, integrate with official government systems, and serve as a template for civic engagement platforms across Africa. We are not just building an app, we are **engineering a more transparent, responsive, and resilient democracy**.

Prepared by: WanaIQ Development Team (Nahashon Mwangi)

Submission Date: January 30, 2026

Contact: Via Hackathon Portal

Document Version: 1.5 (Final Submission)
